Agr. Nat. Resour. 56 (2022) 725–736

AGRICULTURE AND NATURAL RESOURCES

Journal homepage: http://anres.kasetsart.org

Research article

Deep convolutional neural networks with transfer learning for estimation of Nile Tilapia mass

Poonpat Poonnoy[a], **Nissara Kitcharoen**[b,*]

[a] *Faculty of Engineering and Agro-Industry, Maejo University, Chiang Mai 50290, Thailand*

[b] *Faculty of Fisheries Technology and Aquatic Resources, Maejo University, Chiang Mai 50290, Thailand*

Article Info

Article history:
Received 11 October 2021
Revised 10 March 2022
Accepted 15 May 2022
Available online 26 Aug 2022

Keywords:
Computer vision,
Deep learning,
Mass estimation,
Nile Tilapia,
Transfer learning

Abstract

Importance of the work: Fish mass is one of the important traits for selective breeding. Acquiring the mass of individual live fish directly using a digital balance is sensitive to vibration due to respiratory system activity by the fish being measured. This time-consuming and labor-intensive process creates adverse effects in sensitive and prone-to-stress fish.

Objectives: To develop deep convolutional neural networks (ConvNets) with transfer learning to estimate the mass of Nile Tilapia from the image of the fish.

Materials & Methods: In total, 3,832 images were captured and individually paired; mass values were used to create the dataset. The dataset was divided into three groups: training, validating and testing. Some state-of-the-art ConvNets, such as AlexNet, GoogLeNet, VGG-16, VGG-19, Inceptionv3, InceptionResNetV2, NASNetMobile and NASNetlarge, were modified to estimate the mass of the samples.

Results: The modified VGG-19 model provided the lowest values for the root mean square error (3.59 g), mean absolute error (2.27 g), mean relative error (0.05%), mean absolute percentage error (4.09%) and coefficient of determination (0.99). However, the reflection from water film in the background had a negative impact on the mass estimation. The processing time per image on the central processing unit and the graphics processing unit were 0.177 s and 0.053 s, respectively.

Main finding: The modified VGG-19 model was suitable for real-time mass estimation of Nile Tilapia for timely selection of the best broodstocks.

* Corresponding author.
E-mail address: nissara@mju.ac.th (N. Kitcharoen)

https://doi.org/10.34044/j.anres.2022.56.4.07

Introduction

Fish mass is one of the morphological traits that is crucial for optimum feed management (Kubitza and Lovshin 1999; Zhang et al. 2020b), oxygen consumption calculation, antibiotic prescription, fish welfare improvement, decisions on grading, and harvesting (Viazzi et al., 2015; Saberioon and Císař, 2018). The mass is also the major morphological trait for selective breeding (Costa et al., 2013). In this context, precise mass value is required and is usually measured using a digital balance. The direct mass measurement provides accurate results, but it is sensitive to vibration due to the respiratory system activity of the live fish being weighed. Residual water on the fish surface or the platform also confounds the measurement; therefore, a water removal process must be applied before placing the fish on the platform. This process prolongs the resident time for mass acquisition. Performing additional data retrieval from each fish, such as identification number and gender further extends the out-of-water duration. This time-consuming (Shi et al., 2020) and labor-intensive process (Viazzi et al., 2015) creates adverse effects on sensitive and prone-to-stress fish (Pickering and Christie 1981; Maule et al., 1989; Zion, 2012; Fernandes et al., 2020), such as nerve excitement, loss of appetite resulting in growth retardation or death (Karplus et al., 2003). Reducing the data acquisition time decreases these adverse effects on the fish.

An image analysis technique has been introduced for mass estimation of the combined mass of numerous fish, such as for Alaskan Pollock (Balaban et al., 2010a), Alaskan salmon (Balaban et al., 2010b), rainbow trout (Gümüş and Balaban, 2010), seabass (Costa et al., 2013), crucian carp (Zhang et al., 2020b) and Nile tilapia (Fernandes et al., 2020). Estimation of fish mass based on image analysis techniques requires at least two steps: feature extraction and conversion of the one or more extracted features to mass value. Multiple shape features, such as area, width and length, have been used as the inputs to a mathematical model for mass estimation (Gümüş and Balaban, 2010; Zion, 2012; Konovalov et al., 2018; Saberioon and Císař, 2018). The mathematical models used to convert the extracted shape features may be simple linear or non-linear regression models or a sophisticated artificial neural network (ANN) model. Balaban et al. (2010a) developed different simple equations for Alaskan Pollock mass estimation based on fish area. A simple power curve model was found to be the best model for the mass estimation of Alaskan salmon as well (Balaban et al., 2010b). Removing the fins and tail did not improve the accuracy for mass estimation of the Alaskan Pollock. The coefficient of determination (R^2) was 0.99. Besides the area, the mass was well estimated using a length-mass function. Venerus et al. (2016) used a power curve model to estimate the mass of *Sebastes oculatus* Valenciennes, 1833 and *Pinguipes brasilianus* Cuvier, 1829. The estimated results paralleled the direct measured mass with an R^2 value of 0.99. Zhang et al. (2020b) developed an image analysis technique to obtain the area, perimeter, length, width, equivalent diameter, Heywood circularity factor, solidity, arc and complexity of a crucial carp fish. These parameters were used as inputs for the ANN model that was trained to estimate the mass. The ANN model provided estimation results with a mean absolute error (MAE) of 9.1 g, a root mean square error (RMSE) of 11.4 g, and an R^2 value of 0.9391. This approach required no removal of the tail and fins from the fish body but used advanced mathematical methods to select appropriate shape features for the ANN model. In contrast, Fernandes et al. (2020) applied a deep convolutional neural network (ConvNet) called 'SegNet' as proposed by Badrinarayanan et al. (2017) to segment the body of a fish (Nile Tilapia) from its fins and the background before determining the body area, length, height and eccentricity because the fins and tail have different densities compared to the body but significantly influence the fish area (Viazzi et al., 2015). These parameters were used as inputs for the multiple regression model that was developed to calculate the mass of the Nile Tilapia body and carcass. The R^2 values for the body and carcass mass estimation using such a method were 0.96 and 0.95, respectively. This method fulfilled the need for estimating fish mass with or without excluding the fins and tails (Balaban et al., 2010a). However, notably, these techniques were complicated because the fish features had to be extracted and used as predictors in the selected mathematical models. State-of-the-art artificial intelligence, namely a deep convolutional neural network with a transfer learning technique, permits the direct conversion of an image to a fish mass without explicit programming.

Lecun et al. (2015) explained that deep ConvNet recognizes images by transforming the raw input image into various degrees of representations needed for detection or classification. Therefore, it is not necessary to perform feature extraction from the images because the features are automatically discovered during the training process during which the parameters of the convolutional ('conv') and the pooling ('pool') layers are tuned (Kaur and Gandhi, 2020). Assigning a fully connected layer and regression output layer at the end of the network creates deep ConvNet for regression.

These final layers convert representations from the previous layer to desired numerical outputs. The performance of the networks depends on the model complexity of the networks; however, there is not a linear relationship between model complexity and accuracy (Bianco et al., 2018). Currently, more than 40 state-of-the-art deep ConvNets have been trained using one million images (ImageNet) during the ImageNet Large-Scale Visual Recognition Competition (imageNet-1k) (Russakovsky et al., 2015). These pre-trained deep ConvNets can be modified and fine-tuned to estimate fish mass. This method is known as 'transfer learning'.

Transfer learning accelerates the progress in deep ConvNets development because the model architecture was properly designed and learnable parameters were well-initialized. Taheri-Garavand et al. (2020) modified VGG-16 architecture to classify common carp fish (*Cyprinus carpio*) regarding freshness. The classification accuracy was found 98.21%. Zhang et al. (2020a) adapted InceptionResNetV2 (Szegedy et al., 2017), Inception-V3 (Szegedy et al., 2016), and ResNet50 (He et al., 2016) to classify sugarcane sugar crystal particles. It was found that the InceptionResNetV2 produced the highest accuracy of 90.1%. Therefore, deep ConvNets can be applied for the estimation of fish mass.

To the best of the authors' knowledge, there has been no report on the estimation of fish mass using transfer learning. Therefore, the current research article investigated the accuracy of exceptional pre-trained deep ConvNets for Nile Tilapia mass estimation.

Materials and Methods

Fish breeding and grow-out

In total, 3,832 fish in the selective breeding program were collected from a commercial organic farm in Saraphi district (1st generation; 2,157 fish) and the Faculty of Fisheries Technology and Aquatic Resources, Maejo University (2nd generation; 1,675 fish), Chiang Mai province, Thailand.

Initially for the 1st generation, 55 male and 220 female brooders were stocked in a mating cage (2 m wide by 2 m long and 1 m deep) at a ratio of 1 male to 4 females for natural mating. After 3 wk of stocking, eggs were collected from the mouths of female breeders every 7 d and transferred to incubating trays where they were incubated by the family for a week until hatching. After that, 300–500 swim-up fry from each of 100 families were stocked in a hapa (1 m wide by 1 m long and 1 m deep). The fry were fed three times daily with 34% protein organic pelleted feed until tagging. Before tagging, fish were anesthetized with clove oil at 60 parts per million (ppm). Then, each fish was then individually implanted in the peritoneal cavity with a passive integrated transponder (PIT) tag. The tagged fish were reared in a hapa (2 m wide by 2 m long and 2 m deep) and fed twice daily with 34% protein organic pelleted feed for 6 wk. Then, a random selection of tagged fish from each family was divided into two groups. Each group was stocked in four separate cages. Each cage was 2 m wide by 2 m long and 1 m deep. The first group was fed with organic pellets (34% protein feed), while the second group was fed with duckweed. The feeding frequency for both groups was twice daily. After 2 mth, the fish were collected from the cages and anesthetized with clove oil (60–100 ppm). Each anesthetized fish was gently transferred to the data acquisition system developed by the authors. After that, the anesthetized fish were released in the cages.

For the 2nd generation, the selected brooder 200 females and 50 males by top Estimated Breeding Value (EBV) selection were stocked in a mating cage (2 m wide by 2 m long and 1 m deep) at a ratio of 1 to male to 4 females for natural mating. The control group consisted of randomly collected 200 females and 50 males. The fish were stocked in a mating cage (4 m wide by 6 m long and 1 m deep) at a ratio of 100 male to 25 females for natural mating. Similar to the 1st generation, 500 swim-up fry from each of 66 families of selected line and 66 families of control line were stocked in a hapa (1 m wide by 1 m long and 1 m deep) until tagging. The tagged fish were then randomly selected from each family and divided into two groups. Each group was stocked in four separate cages. The first group was reared using the Biofloc system, while the second group was reared using an earthen pond system. The feeding frequency for both groups was twice daily. After 4 mth, the fish were collected from the cages and anesthetized with clove oil (60–100 ppm). Each anesthetized fish was gently transferred to the data acquisition system developed by the authors. After that, the anesthetized fish were then released in the cages. The vitality rate of the fish was observed.

All protocols were approved by the Maejo University Animal Care and Use Committee and complied with the animal ethics for scientific purposes of the National Research Council of Thailand (Approval ID: MACUC 0105/2560).

Image and data acquisition system design

A lightbox (50 cm wide by 50 cm long and 80 cm deep) was built to prevent interference from ambient light and to support the data acquisition components (camera, light source and digital balance), as shown in Fig. 1.

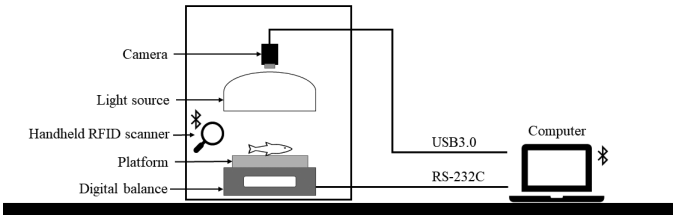

Fig. 1 Schematic diagram of image and data acquisition system, where RFID = radio-frequency identifier and USB = universal service bus

A support structure for the camera and light source was slidable allowing a user to adjust the distance between the camera and the sample. A CMOS camera (BASLER acA2500–14uc; Germany) with a resolution of 2,590 pixels × 1,942 pixels was equipped with an 8.5-mm fixed focal length lens (Edmund Optics; Germany) and connected to a computer via a universal service bus (USB) 3.0 port. The camera was calibrated using a dot grid (4.0 mm in diameter and 9.5 mm spacing). A digital balance (Sartorius: BSA3202S–CW; Germany) with a precision of 0.01 g and a maximum load capacity of 3.2 kg was placed perpendicular to the camera and connected to a computer through an RS-232C port. A green solid Styrofoam slab (30 cm wide by 30 cm long and 2.5 cm thick) was placed over the digital balance and served as a platform to support a sample as well providing a high-contrast background for image processing. To obtain an identification (ID) number for a sample, a radio-frequency identification (RFID) scanner (Biomark: HPR Lite; China) was connected to a computer via Bluetooth®. The hardware was controlled by the software developed with LABVIEW™ ver.2017 (National Instruments, Texas) under the Windows™ 10 environment. The software: 1) communicated with the camera, digital balance and the RFID scanner; 2) managed the image and data logging process to a storage device; and 3) extracted embedded data from collected images.

To ensure the validity of the data, the observed mass and 13-digit ID were embedded in the image as metadata. Once the fish was properly placed on the platform, the fish ID number, mass and image were collected within approximately 15 s. The platform was wiped with a clean cloth to remove residue water and then the balance was tared before sample emplacement. In total, 3,832 fish images were captured and stored on the hard drive. After that, the images were individually processed to extract the embedded data (fish ID and mass). The obtained data associated with its source (image file name and path) were stored in a data table. This table was exported as a comma-separated values (CSV) file.

Dataset preparation

A dataset for deep ConvNets development was created by importing the CSV file into the MATLAB workspace. Then, a data table was built for deep ConvNets development in the MATLAB™ environment. As the input to deep ConvNets was a fish image and the output was its mass, only the file name (including its path) and mass value associated with the individual fish were retrieved. In total, 3,832 records were sorted in ascending order by mass value. Then, the dataset was systematically and randomly selected and divided into three groups for training [1,916 images (approximately 50%)], validating [958 images (approximately 25%)], and testing [958 images (approximately 25%)]. Table 1 shows the characteristics of the sample images in each group based on the observed mass values.

Table 1 Characteristics of sample images in training, validating and testing groups based on observed mass values

Observed mass range (g)	Number of samples			
	Training	Validating	Testing	Total
0–20.00	444	222	222	888
20.01–40.00	328	164	164	656
40.01–60.00	185	93	92	370
60.01–80.00	217	108	109	434
80.01–100.00	180	90	90	360
100.01–120.00	167	84	83	334
120.01–140.00	128	63	64	255
140.01–160.00	89	45	45	179
160.01–180.00	78	39	38	155
180.01–350.00	100	50	51	201
Total	1,916	958	958	3,832

Modification of pre-trained ConvNets architecture

Eight pre-trained ConvNet models were used: AlexNet (Krizhevsky et al., 2017), GoogLeNet (Szegedy et al., 2015), VGG-16 (Simonyan and Zisserman, 2015), VGG-19 (Simonyan and Zisserman, 2015), InceptionV3 (Szegedy et al., 2016), InceptionResNetV2 (Szegedy et al., 2017), NASNetMobile and NASNetLarge (Zoph et al., 2018). These were imported into the MATLAB workspace and modified so that regression analysis could be performed. The modification details were:

1. For Alexnet, VGG-16 and VGG-19, the last three layers of the pre-trained network were replaced with a 'fully connected layer' and a 'regression output layer'. The 'fully connected layer' was connected to the last remaining layer ('drop7') of the pre-trained networks.

2. For GoogLeNet, the last three layers of the network, namely 'loss3–classifier,' 'prob,' and 'output' were replaced by a 'fully connected layer' and a 'regression output layer'. The 'fully connected layer' was connected to the last transferred layer remaining in the network ('pool5–drop_7×7_s1').

3. For InceptionV3 and InceptionResNetV2, the last three layers ('predictions', 'predictions_softmax' and 'ClassificationLayer_predictions') of the network, were substituted with a 'fully connected layer' and a 'regression output layer'. The 'fully connected layer' was connected to the last transferred layer remaining in the network ('avg_pool').

4. For NASNetMobile and NASNetLarge, the last three layers ('predictions', 'predictions_softmax' and 'ClassificationLayer_predictions') of the network were replaced by a 'fully connected layer,' and a 'regression output layer'. The 'fully connected layer' was connected to the 'global_average_pooling2d_1' layer.

It was noted that the substituted 'fully connected layer' had one output. The data from the 'fully connected layer' passed to the 'regression output layer' which was the output of the modified ConvNets.

Training modified ConvNets

The transfer learning ConvNet models were trained with adaptive moment estimation (adam) with a minibatch size of 16. The training options were set as: 'InitialLearnRate' = 0.0001, 'LearnRateSchedule' = 'piecewise', 'LearnRateDropPeriod' = 10, 'LearnRateDropFactor' = 0.1, and 'L2Regularization' = 0.0005. The training process stopped when it reached a maximum epoch of 500 or the RMSE of the validation data set was larger than or equal to the previously smallest RMSE for 10 times. The RMSE was calculated using the formula shown in Equation 1:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\tilde{m}_i - m_i)^2} \quad (1)$$

where m is the observed mass value, $\tilde{m}$ is the predicted mass value and N is the number of samples.

Each image was resized to match the input layer of the pre-trained ConvNets. A workstation (Intel® XEON® Silver 4108 CPU@1.8GHz, RAM 32 GB equipped with NVIDIA Quadro RTX5000) was used in this research.

Statistical analysis

The performance of ConvNets and the regression models were evaluated using the testing dataset. The mean absolute error (MAE), mean relative error (MRE), mean absolute percentage error (MAPE) and coefficient of determination (R^2) were calculated using Equations 2–5, respectively:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\tilde{m}_i - m_i| \quad (2)$$

$$\mathrm{MRE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\tilde{m}_i - m_i}{m_i} \times 100\right) \quad (3)$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left(\left|\frac{\tilde{m}_i - m_i}{m_i}\right| \times 100\right) \quad (4)$$

$$\mathrm{R}^2 = \frac{\left[\sum_{i=1}^{N}(m_i - \bar{m}_i)(\tilde{m}_i - \hat{m}_i)\right]^2}{\sum_{i=1}^{N}(m_i - \bar{m}_i)^2 \sum_{i=1}^{N}(\tilde{m}_i - \hat{m}_i)^2} \quad (5)$$

where m is the observed mass value, $\tilde{m}$ is the predicted mass value, $\bar{m}$ is the mean value of the observed mass values, $\hat{m}$ is the mean value of the predicted mass values and N is the number of samples.

The ConvNet providing the lowest values for MAE, MRE and MAPE was considered as the best model for Nile Tilapia mass estimation.

Results and Discussion

In comparison with the traditional method for fish data acquisition which required transferring fish to multiple measuring stations, the integrated image and data acquisition system developed by the authors reduced the vitality from 10% to 0.1% (data not shown). This was because the data acquisition process was relatively short (15 s) compared to the traditional method (3–5 mins). However, it was observed that the mass acquisition time using the developed system was extended for highly active samples

because the computer could retrieve data only when the platform was stable. This process was conducted in a batch manner with a maximum of 60 samples/hr. For a selective breeding program requiring thousands of samples, the data acquisition should be performed in a continuous manner and be tolerant to vibration caused by the respiratory activity of the sampled fish.

The modified pre-trained ConvNets were modified, trained, validated and tested with the given datasets to estimate the mass of each fish from its image. The loss of these ConvNets decreased as the training process proceeded (Fig. 2). Boxplots and histograms of the relative error (RE) in mass estimation obtained from the modified ConvNets models based on the training datasets are shown in Fig. 3. The modified Alexnet, VGG-16, VGG-19 and GoogLeNet provided more accurate results compared to those from the modified inceptionResNetV2, inceptionV3, NASNetMobile and NASNetLarge. The training time for each model was varied from 26 to 1,679 mins. The modified Alexnet, VGG-16, VGG-19 and GoogLeNet were trained for 26, 33, 154 and 119 mins, respectively. The training processes for these ConvNets were lower than those for the modified inceptionResNetV2 (792 mins), inceptionV3 (375 mins), NASNetMobile (376 mins) and NASNetLarge (1,679 mins). It was noted that the training time increased with the number of layers, connections and learnable parameters (Table 2).

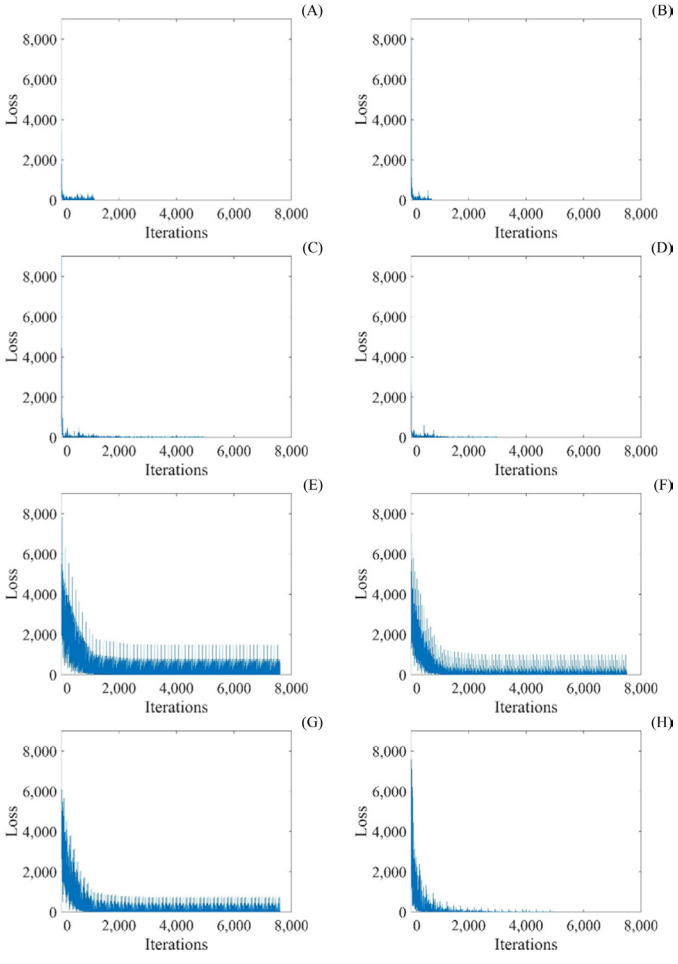

Fig. 2 Loss rate of modified ConvNets models during training process: (A) Alexnet; (B) VGG-16; (C) VGG-19; (D) GoogLeNet; (E) inceptionResNetV2; (F) inceptionV3; (G) NASNetMobile; (H) NASNetLarge

The performance of the modified pre-trained ConvNets on the testing dataset is shown in Table 2. Among the serial networks (AlexNet, VGG-16 and VGG-19), VGG-19 provided the lowest values for RMSE of 3.59 g, MAE of 2.27 g, MRE of 0.05%, MAPE of 4.09% and R^2 of 0.99. However, the errors for mass estimation obtained from the modified directed acyclic graph (DAG) networks were higher than those for the serial networks. The modified GoogLeNet provided the lowest RMSE of 3.92 g, MAE of 2.54 g, MRE of 0.12%, MAPE of 4.70% and R^2 of 0.99. The boxplots and histograms of the estimation RE for the modified ConvNets based on the testing datasets are shown in Fig. 4.

Based on these results, it was noted that the structure of ConvNets influenced the network performance. The total number of connections in the serial networks are not comparable to those of the DAG networks but their convolutional layers effectively extract essential features for mass estimation. VGG-19 (Fig. 5) is deeper than AlexNet (Fig. 6); therefore, it has a better feature learning ability than AlexNet. VGG-19 increased the number of convolutional layers from 8 in AlexNet to 16 layers. The very small filters with 3 pixels × 3 pixels were used in all convolutional layers through the network. These filters activated details in the neighboring pixels (Vizcarra et al., 2021). The combination of a 3 × 3 convolution layer and a 2 × 2 pooling layer in every convolutional block of VGG-19 improved the generalization capability compared to the models with complicated structures (Guan et al., 2019). The models with more layers and connections in the modified DAG networks such as GoogLeNet performed more complex transformations than those required for Nile Tilapia mass estimation; consequently, higher errors were obtained. A similar result was reported for the development of a classification procedure for leaf images (Vizcarra et al., 2021).

Fig. 7. illustrates the maximum activations from different convolutional layers on VGG-19. In the very first convolutional layer, the 'conv1_1' layer emphasized the pattern on the fish. Further, the 'conv1_2' layer separated the fish from the background. The Nile Tilapia image was transformed into an abstract level as it passed through the deeper convolutional layers, as shown in Fig. 8. The output from channel 42 of the last convolutional layer ('conv5_4') highlighted the fish.

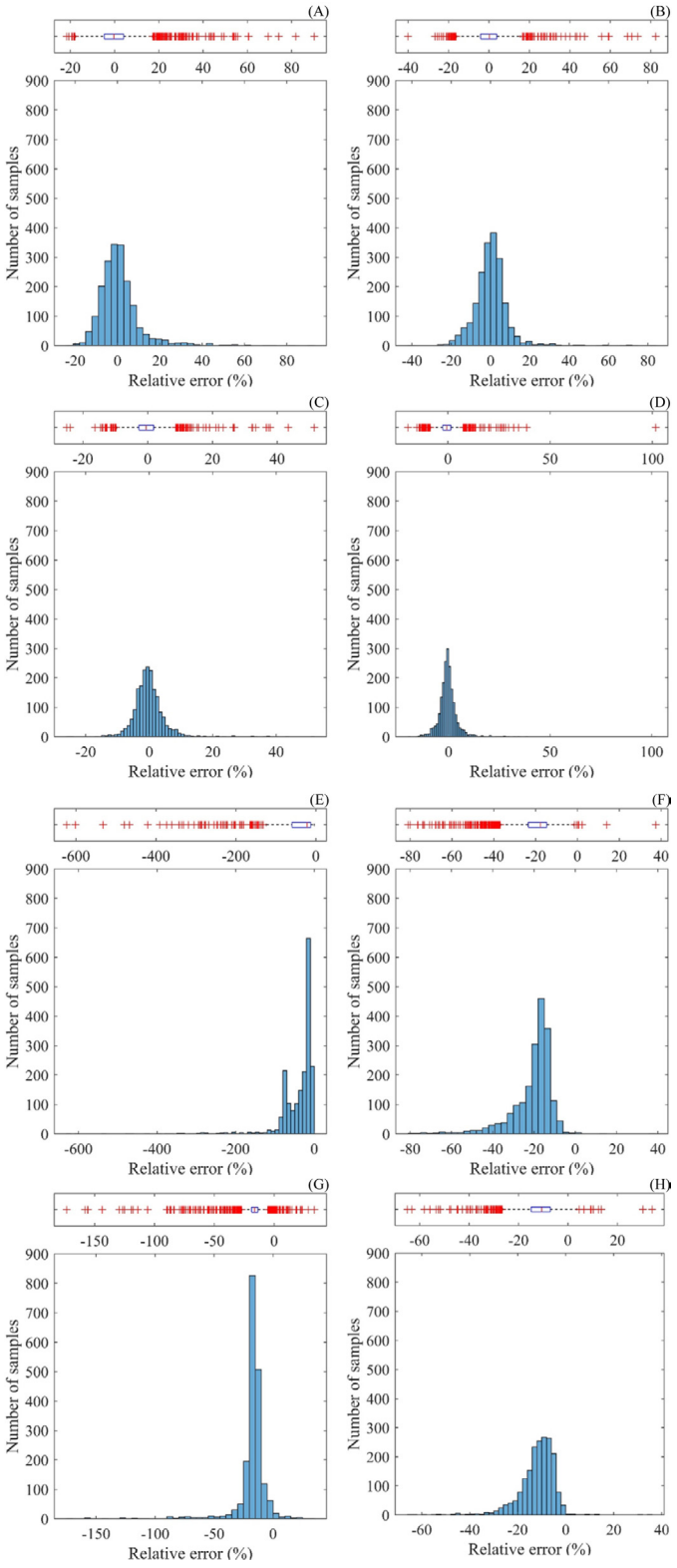

Fig. 3 Boxplots and histograms of relative error from modified ConvNets models based on the training datasets: (A) Alexnet; (B) VGG-16; (C) VGG-19; (D) GoogLeNet; (E) inceptionResNetV2; (F) inceptionV3; (G) NASNetMobile; (H) NASNetLarge

Table 2 Performance and some properties of transfer learning convolutional neural networks for estimation of Tilapia mass

ConvNets	RMSE (g)	MAE (g)	MRE (%)	MAPE (%)	R^2	GPU time (s)	CPU time (s)	Layers	Connections	Learnable parameters (millions)	Size (MB)
AlexNet*	5.19	3.53	1.54	7.40	0.99	0.050	0.059	24	23	56.87	207.2
VGG-16*	5.61	3.47	0.54	6.14	0.99	0.052	0.158	40	39	134.26	488.0
VGG-19*	3.59	2.27	0.05	4.09	0.99	0.053	0.177	46	45	139.57	507.4
GoogLeNet**	3.92	2.54	0.12	4.70	0.99	0.047	0.070	143	169	5.97	22.2
InceptionV3**	14.13	11.78	-21.07	21.08	0.99	0.054	0.153	314	348	21.79	79.4
InceptionResNET**	14.36	13.39	-39.01	39.01	0.99	0.062	0.783	823	920	54.06	198.3
NASNetMobile**	8.13	6.36	-11.40	18.07	0.99	0.085	0.160	912	1071	4.27	15.8
NASNetLarge**	14.47	11.39	-17.63	11.66	0.99	0.056	1.063	1242	1461	88.90	309.5

* = serial network; ** = directed acyclic graph network; RMSE = root mean square error; MAE = mean absolute error; MRE = mean relative error; MAPE = mean absolute percentage error; R^2 = coefficient of determination; GPU = graphics processing unit; CPU = central processing unit

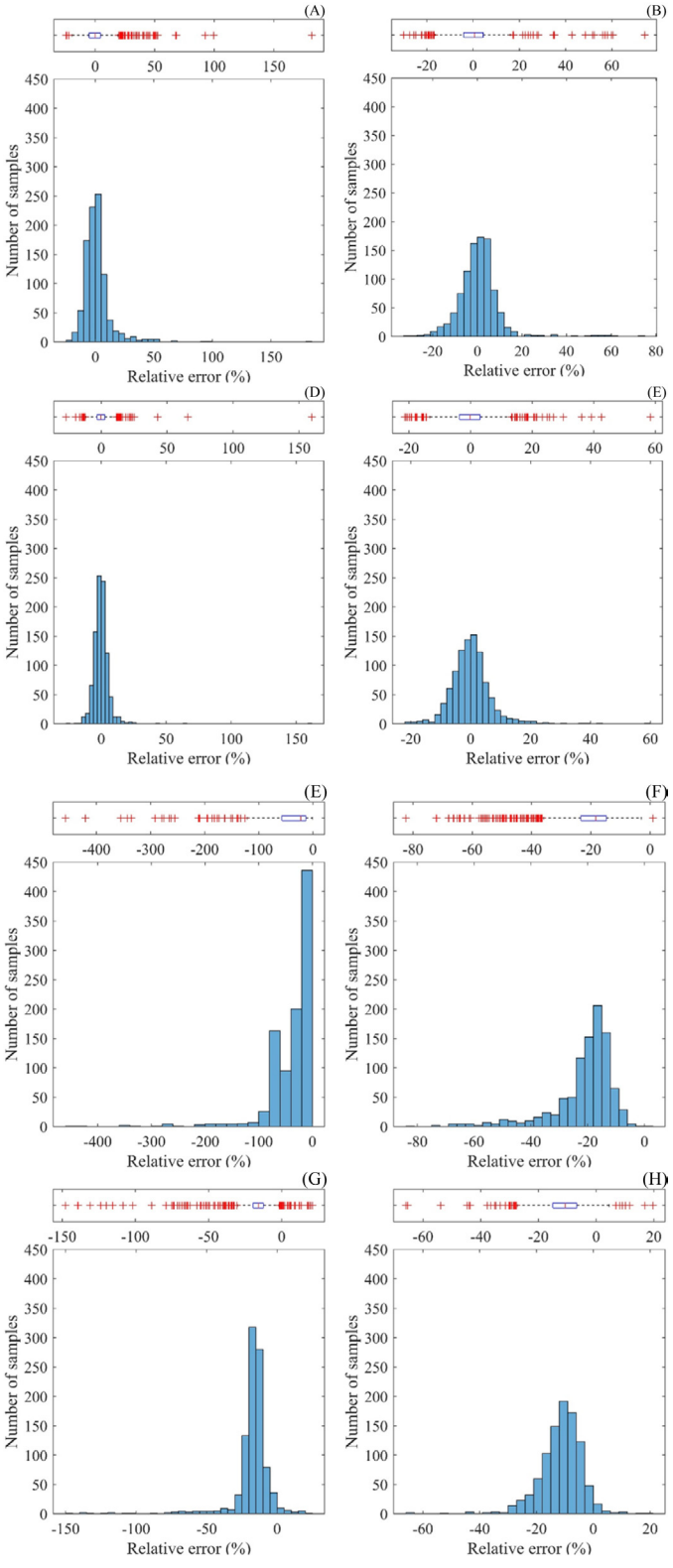

Fig. 4 Boxplots and histograms of relative error from modified ConvNets models based on testing datasets: (A) Alexnet; (B) VGG-16; (C) VGG-19; (D) GoogLeNet; (E) inceptionResNetV2; (F) inceptionV3; (G) NASNetMobile; (H) NASNetLarge

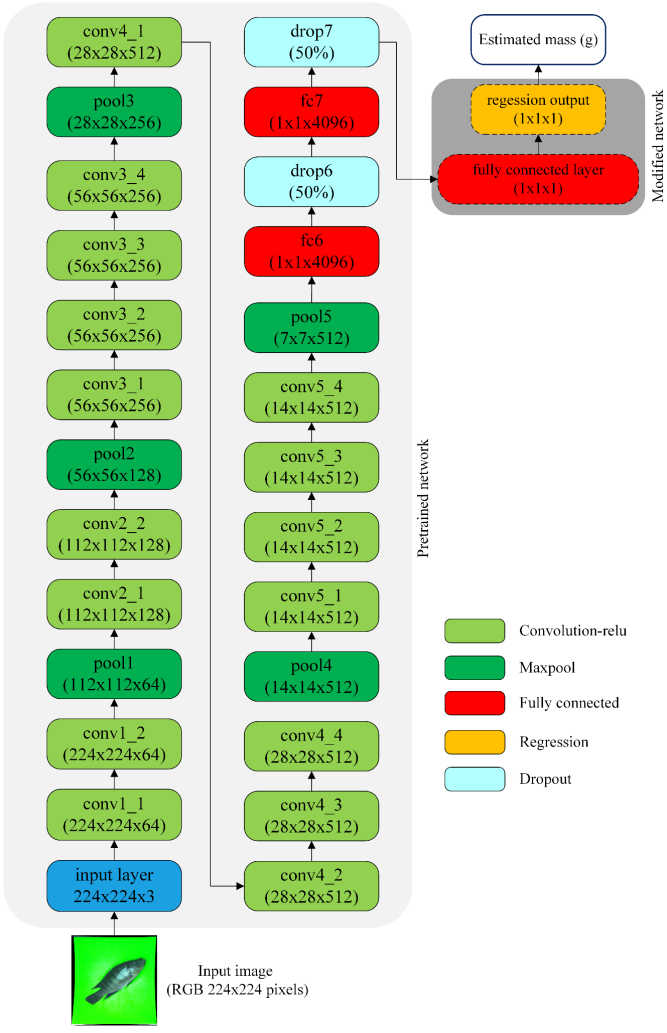

Fig. 5 Schematic diagram of modified VGG-19

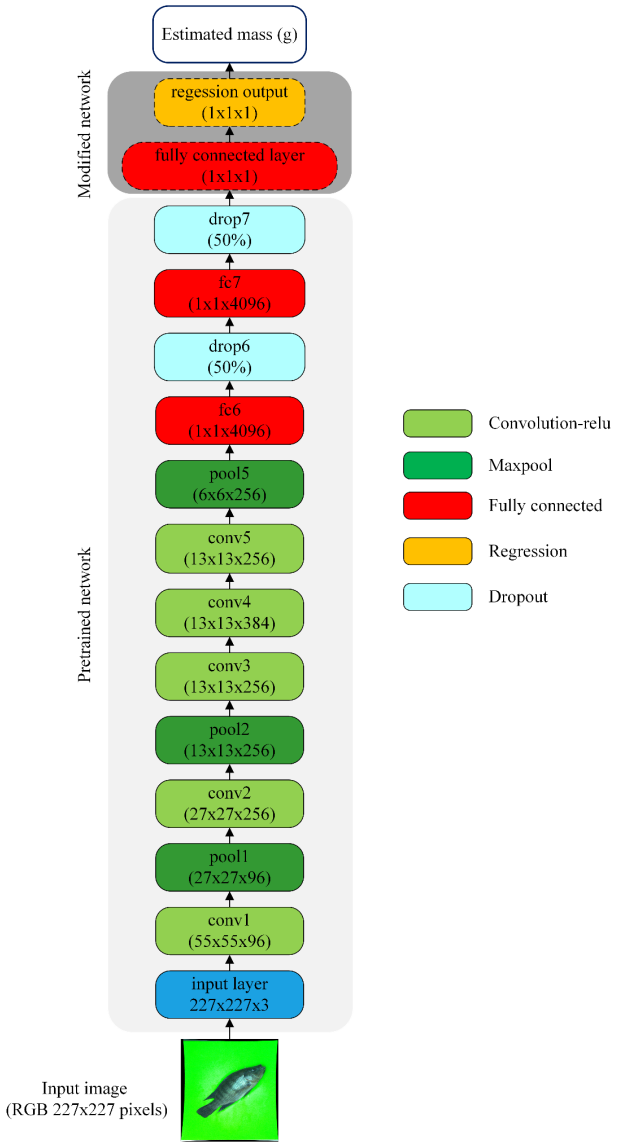

Fig. 6 Schematic diagram of modified AlexNET

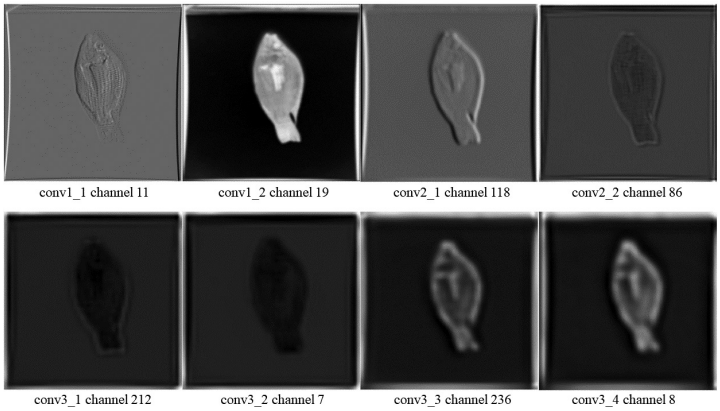

Fig. 7 Maximum activations from 1st–8th convolutional layers of modified VGG-19 network

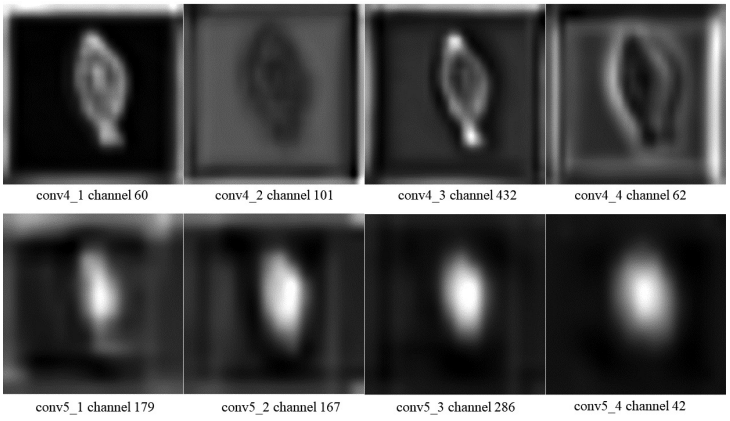

Fig. 8 Maximum activations from 8th–16th convolutional layers of modified VGG-19 network

It was noted that the modified VGG-19 inaccurately estimated the mass of some fish when reflection on the background existed. The water film on the fish body and the background created a strong reflection on the camera causing a white spot on the image. The maximum activations of this image were different from the input image without reflection on the image. In this case, the result from the modified VGG-19 model (10.38 g) was higher than the observed value (6.08 g). Upgrading the image acquisition system by adding a polarizing filter to the camera lens may reduce the reflection and overcome the limitation of mass estimation from the fish image using the proposed deep ConvNets with transfer learning.

The plots between the observed and estimated mass from the modified VGG-19 model on the testing dataset ($n = 938$) are shown in Fig. 9. The estimated mass values of male and female fish in the first (blue dots) and second (red dots) generations paralleled the observed ones ($R^2 = 0.99$). The results indicated that the relationship between the appearance and the mass of each fish varied by age, rearing condition, sex, feed and genetics (Fernandes et al., 2015) and these associations was recognized by the modified VGG-19.

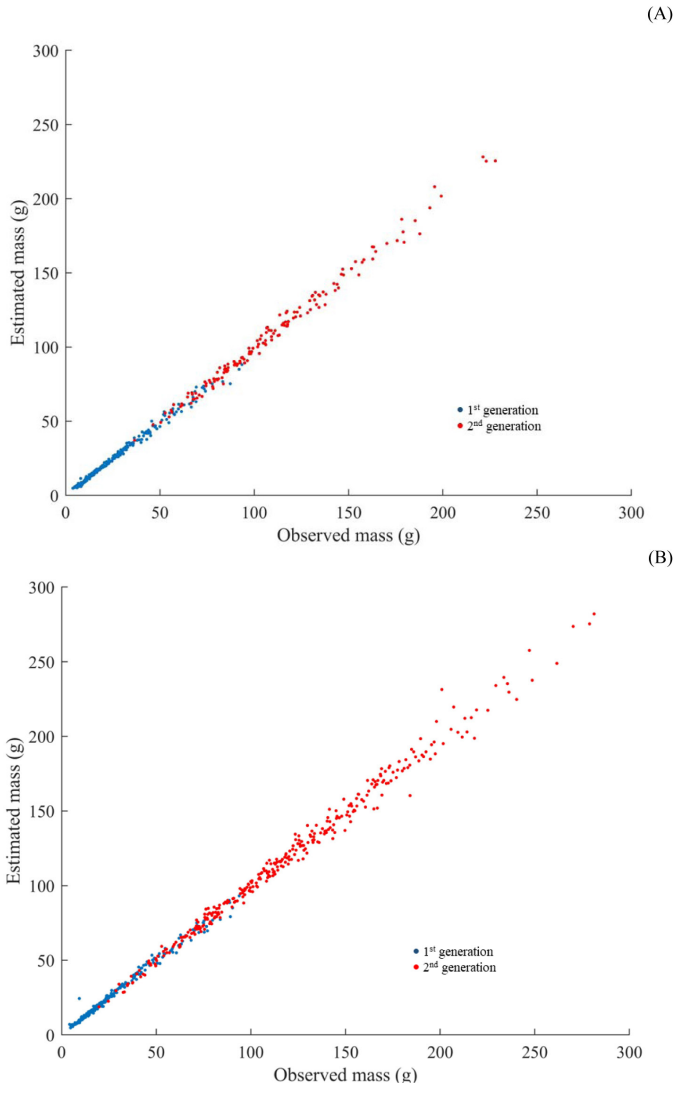

Fig. 9 Observed and estimated mass values using modified VGG-19 model based on testing dataset (n = 958): (A) female fish; (B) male fish

Additional experiments on mathematical model development using the traditional least-squares method were carried out to estimate Nile Tilapia mass from shape features (total area or total length) obtained by image processing (details not shown). The performance levels from mass estimation using the testing dataset were determined and compared. It was found that the power curve model using area as input provided the lowest RMSE 18.60 g, MAE of 2.69 g, MRE of 2.19% and MAPE of 4.50%. The modified VGG-19 was superior to the power curve model. Fernandes et al. (2020) suggested the use of body area for estimation of Nile Tilapia mass. This method required SegNet to remove the fins and tail from the body. The body area-based model provided a higher RMSE (77.52 g), MAE (57.77 g) and MAPE (11.35%). It was noted that the average mass of fish samples used in such a study was 523.8 g (SD = 224.4) which was different from the present study of 72.2 g (SD = 57.66). The performance of the VGG-19 model on mass estimation of large Nile Tilapia should be further investigated.

The proposed modified VGG-19 model required an RGB image with a resolution of 224 pixels by 224 pixels and provided superior mass estimation accuracy without manual features extraction. The processing time per image for the graphics processing unit (GPU) and central processing unit (CPU) for estimation of the Nile Tilapia mass using the modified VGG-19 method were 0.053 s and 0.177 s, respectively. Therefore, the fish mass could be obtained at a maximum of 1,132 samples/min. The modified VGG-19 could be strongly recommended for real-time mass estimation of Nile Tilapia. Further development of the conveying system for continuous image acquisition is necessary. The mass estimation performance could be improved by using 3D imaging because then the thickness of fish would be taken into account (Konovalov et al., 2018).

Conclusion

Fish mass is one of the major parameters used in selective breeding and aquaculture management. VGG-19 (a state-of-the-art deep convolutional neural network) was modified to estimate the mass of a Nile Tilapia fish directly from its two-dimensional image. The estimated mass was consistent with the observed mass. The processing time on the CPU and GPU of the modified VGG-19 model were less than one second; therefore, it can be applied for real-time mass estimation of Nile Tilapia. Decreasing the time for fish mass acquisition reduces the data acquisition time and labor, as well as reducing any adverse health effects to fish. For a fish selective breeding program, thousands of individual fish must be measured; thus, application of the proposed method may accelerate the data acquisition process and reduce losses of brooders. Computer vision combined with a deep learning technique could possibly be applied for mass estimation of other fish species. However, development of the deep convolutional neural networks requires a large dataset of fish images and the mass value of each fish. Additionally, the fish images must be taken under controlled conditions to maintain the image proportions while also maintaining uniform light distribution and minimizing ambient light interference. Reflection of water film in the background should be minimized to maintain the optimum

estimation performance conditions. Using three-dimensional imaging may increase the performance in mass estimation as the thickness of fish could be incorporated.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Acknowledgements

Muang-Jai Organic Farm, Chiang Mai, Thailand provided valuable cooperation throughout the research. Ms. Puncharat Meekaew, Ms. Kanokwan Nakkham and Mr.Supakrit Khanjungreed assisted during sample preparation. The students and staff of the Faculty of Fisheries Technology and Aquatic Resources, Maejo University provided support during the experimental period.

References

Badrinarayanan, V., Kendall, A., Cipolla, R. 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39: 2481–2495.

Balaban, M.O., Chombeau, M., Cırban, D, Gümüş, B. 2010a. Prediction of the weight of Alaskan pollock using image analysis. J. Food Sci. 75: E552–556. doi.org/10.1111/j.1750-3841.2010.01813.x

Balaban, M.O., Unal Sengör, G.F., Soriano, M.G., Ruiz, E.G. 2010b. Using image analysis to predict the weight of Alaskan salmon of different species. J. Food Sci. 75: E157–162. doi.org/10.1111/j.1750-3841.2010.01522.x

Bianco, S., Cadene, R., Celona, L., Napoletano, P. 2018. Benchmark analysis of representative deep neural network architectures. IEEE Access 6: 64270–64277. doi.org/10.1109/ACCESS.2018.2877890

Costa, C., Antonucci, F., Boglione, C., Menesatti, P., Vandeputte, M., Chatain, B. 2013. Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis. Aquacult. Eng. 52: 58–64. doi.org/10.1016/j.aquaeng.2012.09.001

Fernandes, A.F.A., Silva, M.d.A., Alvarenga, E.R.d., et al. 2015. Morphometric traits as selection criteria for carcass yield and body weight in Nile tilapia (*Oreochromis niloticus* L.) at five ages. Aquaculture 446: 303–309. doi.org/10.1016/j.aquaculture.2015.05.009

Fernandes, A.F.A., Turra, E.M., Alvarenga, É.R.d., Passafaro, T.L., Lopes, F.B., Alves, G.F.O., Singh, V., Rosa, G.J.M. 2020. Deep learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia. Comput. Electron. Agr. 170: 105274. doi.org/10.1016/j.compag.2020.105274

Guan, Q., Yunjun, W., Ping, B., et al. 2019. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. J. Cancer 10: 4876–4882. doi.org/10.7150/jca.28769

Gümüş, B., Balaban, M.O. 2010. Prediction of the weight of aquacultured rainbow trout (*Oncorhynchus mykiss*) by image analysis. J. Aquat. Food Prod. Technol. 19: 227–237. doi.org/10.1080/10498850.2010.508869

He, K., Zhang, X., Ren, S., Sun, J. 2016. Deep residual learning for image recognition. 2016. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, pp.770–778.

Karplus, I., Gottdiener, M., Zion, B. 2003. Guidance of single guppies (*Poecilia reticulata*) to allow sorting by computer vision. Aquacult. Eng. 27: 177–190. doi.org/10.1016/S0144-8609(02)00085-7

Kaur, T., Gandhi, T.K. 2020. Deep convolutional neural networks with transfer learning for automated brain image classification. Mach. Vis. Appl.s 31: 20. doi.org/10.1007/s00138-020-01069-2

Konovalov, D.A., Saleh, A., Domingos, J.A., White, R.D., Jerry, D.R. 2018. Estimating mass of harvested Asian seabass *Lates calcarifer* from images. World J. Eng. 6: 15–23. doi.org/10.4236/wjet.2018.63B003

Krizhevsky, A., Ilya, S., Hinton, G.E. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60: 84–90. doi.org/10.1145/3065386

Kubitza, F., Lovshin, L.L. 1999. Formulated diets, feeding strategies, and cannibalism control during intensive culture of juvenile carnivorous fishes. Rev. Fish. Sci. 7: 1–22. doi.org/10.1080/10641269991319171

LeCun, Y., Bengio, Y., Hinton, G. 2015. Deep learning. Nature 521: 436–444. doi.org/10.1038/nature14539

Maule, A.G., Tripp, R.A., Kaattari, S.L., Schreck, C.B. 1989. Stress alters immune function and disease resistance in chinook salmon (*Oncorhynchus tshawytscha*). J. Endocrinol. 120: 135–142. doi.org/10.1677/joe.0.1200135

Pickering, A.D., Christie, P. 1981. Changes in the concentrations of plasma cortisol and thyroxine during sexual maturation of the hatchery-reared brown trout, *Salmo trutta* L. Gen. Comp. Endocrinol. 44: 487–496. doi.org/10.1016/0016-6480(81)90337-3

Russakovsky, O., Jia, D., Su, H., et al. 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115: 211–252. doi.org/10.1007/s11263-015-0816-y

Saberioon, M., Císař, P. 2018. Automated within tank fish mass estimation using infrared reflection system. Comput. Electron. Agr. 150: 484–492. doi.org/10.1016/j.compag.2018.05.025

Shi, C., Wang, Q., He, X., Zhang, X., Li, D. 2020. An automatic method of fish length estimation using underwater stereo system based on LabVIEW. Comput. Electron. Agr. 173: 105419. doi.org/10.1016/j.compag.2020.105419

Simonyan, K., Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA, pp. 1–14.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alexander, A.A. 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco, CA, USA, pp. 4278–4284.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. 2016. Rethinking the inception architecture for computer vision. 2016. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, pp. 2818–2826.

Szegedy, C., Wei, L., Jia, Y., et al. 2015. Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA, pp. 1–12. doi:10.1109/CVPR.2015.7298594

Taheri-Garavand, A., Nasiri, A., Banan, A., Zhang, Y. 2020. Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish. J. Food Eng. 278: 109930. doi.org/10.1016/j.jfoodeng.2020.109930

Venerus, L.A., Gomila, G.L.V., Sueiro, M.C., Bovcon, N.D. 2016. Length-weight relationships for two abundant rocky reef fishes from northern Patagonia, Argentina: *Sebastes oculatus* Valenciennes, 1833 and *Pinguipes brasilianus* Cuvier, 1829. J. Appl. Ichthyol. 32: 1347–1349. doi.org/10.1111/jai.13207

Viazzi, S., Hoestenberghe, S.V., Goddeeris B.M., Berckmans, D. 2015. Automatic mass estimation of Jade perch *Scortum barcoo* by computer vision. Aquacult. Eng. 64: 42–48. doi.org/10.1016/j.aquaeng.2014.11.003

Vizcarra, G., Bermejo, D., Mauricio, A., Gomez, R.Z., Dianderas, E. 2021. The Peruvian Amazon forestry dataset: A leaf image classification corpus. Ecol. Inform. 62: 101268. doi.org/10.1016/j.ecoinf.2021.101268

Zhang, J., Meng, Y., Wu, J., Qin, J., wang, H, Yao, Y., Yu, S. 2020a. Monitoring sugar crystallization with deep neural networks. J. Food Eng. 280: 109965. doi.org/10.1016/j.jfoodeng.2020.109965

Zhang, L., Wang, J., Duan, Q. 2020b. Estimation for fish mass using image analysis and neural network. Comput Electron Agr. 173: 105439. doi.org/10.1016/j.compag.2020.105439

Zion, B. 2012. The use of computer vision technologies in aquaculture – A review. Comput. Electron. Agr. 88: 125–132. doi.org/10.1016/j.compag.2012.07.010

Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V. 2018. Learning transferable architectures for scalable image recognition. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. doi:10.1109/CVPR.2018.00907